

Variable aggregation for dynamical systems on networks

Mauro Faccin, Jean-Charles Delvenne & Michael Schaub

Topological Data Analysis (TDA) offers methodologies to explore high-dimensional datasets extracting geometrical properties to infer the relevant patterns of data. These methods often require the tuning of parameters which demands either pre-existing knowledge of data or time-consuming trial and error. We present STAD (Simplified Topological Approximation of Data) a new algorithm to approximate the topology of high-dimensional data without requiring additional parameters or formulating prior hypothesis.

STAD generates an abstract representation of high-dimensional data by giving each data point a location in a graph which preserves the distances in the original high-dimensional space. The STAD graph is built upon the Minimum Spanning Tree (MST) to which new edges are added until the correlation between the graph and the original dataset is maximised.

Additionally, STAD supports the inclusion of filter functions to focus the exploration and allow the analysis of data from new perspectives, emphasising traits in data which otherwise would remain hidden.

We illustrate the use of our method by applying it to three different kinds of data such as the traffic density from Barcelona, the temporal measurements of the quality of the air in Castile and León and free text in patient records, in each case finding signals for others classical techniques would be indistinguishable.